M1 蛋白質分子模型建構

楊健志

蛋白質的基本單元是胺基酸,胺基酸的序列稱爲蛋白質的一級結構,在三度空間有秩序的摺疊則構成其三級結構。胺基酸藉由胜肽鍵組成胜肽。胜肽平面與 $C\alpha$ 形成的雙面角 (dihedral angle) 分別爲 ϕ (phi) 角與 ψ (psi) 角。大部分蛋白質中 phi 與 psi 的組合會落在 Ramachdran plot 固定的區域,亦即蛋白質主要的二級結構特徵 α -helix 及 β -sheet。要了解蛋白質結構與功能的關係,最好的方法就是利用預製的 原子組件去組合胺基酸及胜肽。

1. 器材

HGS Biochemistry Molecular Model,5001 Protein Nucleic Acid Set (Maruzen, Japan)。模型中 1 公分相當於 1 Å。請想想,爲何發給你的 單元 (building block) 不是標準的胺基酸。注意,細小的零件要小心照顧。

2. 實驗操作

請先熟悉兩種幫助描述空間中原子相對排列位置的方法。第一,除了 Glycine 以外,所有的胺基酸都有鏡像異構物,分別為 D-form,L-form。通常以 L-glyceraldehyde 為標準,與其有類似原子安排的胺基酸為 L-form。為了方便,我們採用一個口訣 "CO-R-N"。第二,雙面角 (dihedral angle) 描述相連的四個原子的相對位置,亦即前三個原子與後三個原子所構成的兩個平面之間的夾角。 ϕ 角由 C'-N-C $_{\alpha}$ -C' 形成,頭尾兩個原子距離最遠時為 180°,距離最近時為 0°,由 N 往 C $_{\alpha}$ 看,C'-N 所在位置為 0,C'往順時針方向偏為正角。 ψ 角 由 N-C $_{\alpha}$ -C'-N 形成,依上所述類推。操作中請用尺量度二級結構中的鍵長,直徑等特性值,並紀錄之。

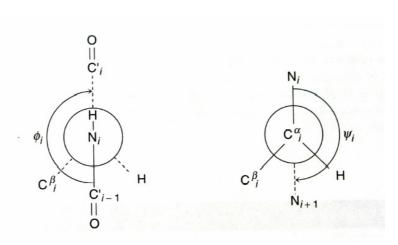
- 2.1. 請小心分開一個"胺基酸"單元,請注意代表原子的顏色。
 - 2.1.1. Red, oxygen; blue, nitrogen; black, carbon; white, hydrogen.
 - 2.1.2. 請辨別, sp3 carbon 爲 Cα; sp2 carbon 爲胜肽鍵的一部分。Hydrogen atom

及 hydrogen bond 也有所不同。

2.2. 請拿出 sp3 carbon,接上 hydrogen atom。母端保留給 R-group (side chain),選擇剩下兩個公端其中之一依 CORN 原則接上 nitrogen,即完成 L-form configuration。



- 2.3. 利用 partial double bond (灰色小管),以 trans form 形式連接 N 與 C'。
- 2.4. 將 hydrogen bond 及 oxygen 接上,檢查完成的單元是否與未分解的單元完 全一致。
- 2.5. 將單元 1 上 Cα 剩餘之公端與單元 2 之 C'相接,再依序接上單元 3,並使 兩相鄰 C'或相鄰的 N 保持最遠距離 (完全延伸)。請觀察你所得到的一個"近似" tripeptide 的構造,它的 N-terminal 在哪裡?
- 2.6. 試著轉動 N-Cα bond,此即 φ角。轉動 Cα-C' bond,此即 ψ角。
- 2.7. 當 tripeptide 爲 "完全延伸" 時, $\phi = \psi = 180^\circ$,由 N 向 $C\alpha$ 觀察,是否如下圖 所見 (from Fersht)。



- 2.8. 由 N 向 C α 觀察,以靠近你的 C=O 爲 0°,轉動 ϕ 角,先使角度歸零,再 逆時針旋轉 45°。
- 2.9. 由 $C\alpha$ 向 C'觀察,以靠近你的 N-H 爲 0°,轉動 ψ . 角,先使角度歸零,再逆 時針旋轉 45°。依序轉動其後的雙面角。
- 2.10. 與其他組員所得相連,你得到什麼二級結構?
- 2.11. 連結 residue i 之 C=O 與 residue i+4 的 N-H 形成 H-bond。
- 2.12. 觀察所得,並以尺量度所有相對距離。

2

- 2.13. 小心分開單元,組成兩條完全延伸的胜肽,使成"反平行"排列,是否可以 找到很接近的 hydrogen bond donor 及 acceptor 使之相接。你得到什麼二級結 構?
- 2.14. 組成兩條完全延伸的胜肽,使成"平行"排列,是否可以找到很接近的 hydrogen bond donor 及 acceptor 使之相接? H-bond

3. 參考文獻:

- 3.1. Anna Tramontano (2006) Protein structure prediction. Wiley-VCH
- 3.2. Höltje et al (2003) Molecular Modeling. Wiley-VCH
- 3.3. Current Protocol in Bioinformatics. Wiley online book.

M2 生物資訊基礎

楊健志

生物資訊起源於蛋白質,及核酸序列的比較,1965 年 Margaret Dayhoff's 根據當時有限的蛋白質分子序列,建立資料庫,並利用電腦及數學探討生物分子間在演化上的關係。生物資訊需應用生物學,遺傳學,生物化學,數學,資訊,統計等領域的知識,以解決生物學上的問題。現代生物資訊的範疇至少包括,(1) 建立新的演算法或統計方法以研究大型資料組之間的關係; (2) 分析不同形式的生物資訊如蛋白質序列,蛋白質功能區塊,或三級結構; (3) 建立有效率的資料庫與分析工具。本實驗利用以自動定序儀完成的核酸定序,解讀其檔案,找出核酸目標序列,轉譯成蛋白質序列,搜尋資料庫,區塊分析,多重序列比對,二級結構預測,三級結構模擬。進行這些工作之前,我們要先了解有哪些可用的生物資訊工具,及不同的資料庫。

1. 資料庫與分析工具

1.1. 資料庫

- 1.1.1. SwissProt (http://www.expasy.ch)
 - 1.1.1.1. 整合型蛋白質一級結構資料庫,包含兩種序列資料庫,Swissprot 及 TrEMBL。另外也含有 ENZYME Database,提供依酵素功能分類的酵素資料,可用 EC number 查找酵素,如 EC 1.1.1.1。資料庫中含有眾多有用的分析工具及外部連結,如 BLAST,ScanProsite,SwissPDB viewer,SwissModel。
- 1.1.2. NCBI (http://www.ncbi.nlm.nih.gov/)
 - 1.1.2.1. 整合型資料庫,提供核酸,蛋白質,文獻,及最近加入的化學分子 資料庫。
- 1.1.3. EMBL (http://www.ebi.ac.uk/embl/)
 - 1.1.3.1. 核酸序列資料庫。由 EBI 維護, 在 Sanger centre 的隔壁。
- 1.1.4. PDB (Protein Data Bank, http://www.rcsb.org)
 - 1.1.4.1. 蛋白質三級結構資料庫。

1.2. 整合型分析工具

1.2.1. EMBOSS

1.2.1.1.1.1. 開放自由軟體,可以下載其原始碼(http://emboss.sourceforge.net/),提供個人使用或以網路型式提供服務。除了原始的 UNIX 界面,也有 JAVA 介面,及方便的圖型使用者介面 (GUI, graphical user interface)。以"EMBOSS GUI"搜尋 Google 可找到眾多公開已設置的服務網頁,如 http://anabench.bcm.umontreal.ca/html/EMBOSS/。

1.2.2. GCG

1.2.2.1. 以大型伺服主機對用戶提供的整合型資料庫與分析工具軟體,原來 是 UNIX 介面,現在也提供網頁介面。

1.2.3. Vector NTI

商用個人電腦整合型軟體,本系採購四套單機版及一套網路版執照。

2. 常用分析指令或軟體

	GCG	EMBOSS	Comment
輸入定序檔		ABIview	
核酸序列分析			
	CodonFrequency	chips compseq cusp	CodonFrequencytabulates codon usage. compseq counts composition of dimer/trimer in sequence. chips calculates codon usage stats cusp creates a codon usage table.
	Frames	plotorf showorf	Show open reading frames. plotorf does this graphically
	Map Mapplot Mapsort	restrict remap restover	finds restriction enzyme cleavage sites. GCG & EMBOSS may display different isoschizomers of the same enzyme, but the results are equivalent. The EMBOSS remap program may not display a few of the available isoschizomers.
	StemLoop	<u>palindrome</u> <u>etandem</u>	Finds inverted repeats.
	<u>Testcode</u>	wobble	Plots 3rd-position variability as an indicator of potential coding regions.
	<u>Translate</u>	transeq	Translates nucleotide -> Protein sequences
序列搜尋			
	BLAST	dbiBlast	
			SwissProt, NCBI,EMBL
	FASTA	N/A	
	Assemble	merger	Construct new sequences from pieces of existing sequences; merger only accepts 2 sequences while assemble accepts several.
序列比對			

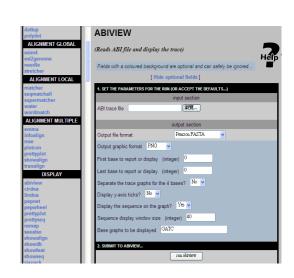
-				
	<u>BestFit</u>	water matcher	Bestfit uses the Smith-Waterman algorithm to find the best local alignment between 2 sequences. water uses Smith-Waterman, matcher uses Pearson's lalign algorithm.	
	Gap	needle stretcher	Needleman-Wunsch algorithm to compare 2 sequences. stretcher uses the Myers-Miller algorithm which is more memory-efficient. water->matcher->supermatcher are local alignment programs for small, medium, and large sequences, respectively.	
	<u>Prime</u>	eprimer3	Selects oligonucleotide primers.	
一級結構分析				
	Composition	compseq pepstats	Sequence composition	
	<u>HelicalWheel</u>	<u>pepwheel</u>	Plots peptide sequence as helical wheel to help recognize amphiphilic regions.	
	HTHScan	helixturnhelix	Finds HTH motifs in protein sequences.	
	<u>Motifs</u>	patmatmotifs	Finds common Prosite motifs in a sequence.	
	<u>Pepplot</u>	pepinfo	Pepplot plots protein 2ndary structure and hydrophobicity, pepinfo plots hydrophobicity, and garnier does protein 2ndary structure prediction.	
	Composition	compseq pepstats	Sequence composition	
		antigenic	Finds antigenic sites in proteins	
		<u>tmap</u>	Displays membrane spanning regions	
二級結構分析	另有許多資料庫網頁 http://www.igb.uci.edu/?page=tools&subPage=psss			
	Peptidestructure Plotstructure	garnier	Secondary structure prediction. Garnier does not include Jameson-Wolf antigenic indexing.	
多重序列比對				
	Pileup	emma	Multiple sequence alignment, emma is an interface to ClustalW. Can also use the standalone Clustal, or web ClustalW.	
其他				
		<u>findkm</u>	Find Km and Vmax for an enzyme reaction by a Hanes/Woolf plot	

3. 實驗操作

請依指示下載定序檔。

3.1. 核酸序列輸出及搜尋

- 3.1.1. 找到可用的 EMBOSS GUI server。
- 3.1.2. 從網頁左方點選 abiview,利用"瀏覽"
- 3.1.3. 輸入之前下載的定序檔 .abi 。
- 3.1.4. 點選 "run abiview"。



- 3.2. 轉譯為蛋白質序列
 - 3.2.1. 從 abiview 輸出的序列,複製 層析圖訊號清楚的片段。
 - 3.2.2. 從網頁左方點選 transeq,貼上片段,試著找出最長的轉譯序列。
- 3.3. 搜尋全長序列,找出同源蛋白。
 - 3.3.1. 到 Expasy (Swissprot) 首頁,點選 BLAST 工具,貼上轉譯序列。仔細看有何參數及選項。
 - 3.3.2. 觀察 BLAST 的結果網頁,相似度最高者是什麼蛋白質? 點選該筆資料。
- 3.4. 一級結構特性分析
 - 3.4.1. 回到 EMBOSS 首頁,依作業問題尋找適當軟體執行。
- 3.5. 多重序列排比
- 3.6. 二級結構預測。3.6.1.1. 利用 EMBOSS 中 garnier 程式,分析全長蛋白質的二級結構。
- 3.7. 搜尋三級結構資料庫。

4. 參考文獻:

- 4.1.1. http://www.ncbi.nlm.nih.gov/Education/BLASTinfo/milestones.html
- 4.1.2. Current Protocols in Bioinformatics, Wiley-VCH
- 4.1.3. Current Protocol in Bioinformatics. Wiley online book.
- 4.1.4. http://helix.nih.gov/apps/bioinfo/emboss-gcg.html

M3 蛋白質分子模擬

楊健志

仔細的研究蛋白質的三級結構增進我們對於功能的了解,例如摺疊形式 (Fold),區塊組合與安排,表面電荷分佈,活性區,基質結合區,分子內交互作用,多元體的分子間交互作用。經實驗方法如 X-ray crystallography,NMR 所決定的蛋白 質 三 級 結 構 , 收 錄 於 蛋 白 質 結 構 資 料 庫 Protein Databank (PDB, http://www.rcsb.org),目前記錄有 42350 個蛋白質結構。PDB 中的蛋白質或其他生物分子的結構是以每一個原子的三度空間座標呈現,有許多公用或自由軟體可用來觀察這些蛋白質三級結構,如 SwissPDBViewer,PyMol,Rasmol等,這樣的工作稱爲分子視算 (Molecular Graphic)。在 SwissProt,NCBI 中收錄有一級結構的蛋白質的數目遠遠超過 PDB 的資料數,這是分子模擬 (Molecular modeling) 發揮應用的地方。根據 Anfinsen 的理論,蛋白質的一級結構決定其三級結構,我們可以合理的延伸這個原理爲,有近似的一級結構的兩種蛋白質,有近似的三級結構,這個假設構成同源分子模擬 (Homology modeling) 的基礎。

同源分子模擬的基本步驟如下。

- (1) 以 BLAST 搜尋結構已知的同源蛋白當模版 (template)。
- (2) 比對 (align) 目標蛋白與模版。
- (3) 根據模版建立模型。
- (4) 根據資料庫模擬 side-chain 及 loop。
- (5) 分子力學及能量最小化。
- (6) 構型分析,檢驗不符蛋白質結構常規者。
- (7) 根據已有功能研究知識,驗證模型,反覆進行模擬。

1. 工具軟體與資料庫

1.1. SwissProt (http://www.expasy.ch)

蛋白質一級結構資料庫,資料完整清楚,歷史悠久,被譽為 Golden Standard。輔助工具眾多,如 SwissModel, SwissPDBviewer。

1.2. PDB (http://www.rcsb.org)

蛋白質三級結構資料庫。

1.3. Discovery Studio 1.7

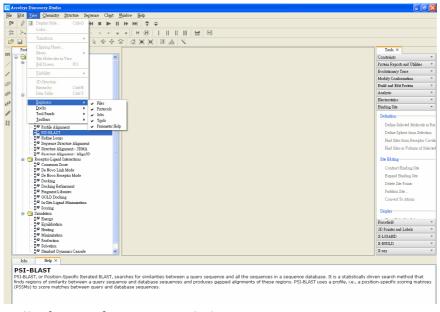
完整的分子模擬軟體套件組,為一商業付費軟體。可做序列分析比對(如BLAST, Multiple sequence alignment),同源分子模擬(Homology modeling),蛋白-配體接合(molecular docking),分子力學(molecular dynamics),藥物設計等。這套軟體中有關分子視覺化的部份,DS visulizer,是免費的。

1.4. Modeller (http://salilab.org/modeller/)

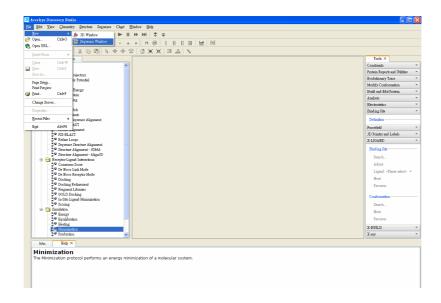
2. 實驗操作

同源分子模擬中已知三級結構的蛋白質稱爲模板 (template),而未知三級結構的稱爲目標 (target)。本實驗將接續上週實驗,利用序列比對的方式,看是否可爲經定序確認的蛋白質找到可用的模板。SwissModel 找到的模板是否與你找到的相同。 β -澱粉酶的摺疊形式是自然界中最常利用的構型,利用 Discovery Studio 爲甘藷中的 β -澱粉酶找到模板,並進行分子模擬,觀察它的結構。並學習如何找到活性區或與基質進行 docking。

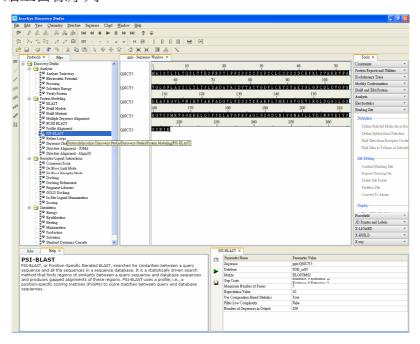
- 2.1. 利用上週的 DNA 定序檔,以 EMBOSS 找到某個蛋白質的部份序列。
- 2.2. 到 Expasy 中搜尋完整序列。
- 2.3. 將這些序列分別送到 SwissModel,是否有結果? 請先使用 First approach mode,並輸入自己的電子郵件帳號。
- 2.4. 到 Expasy 中搜尋甘藷中的 β-澱粉酶完整序列,輸入 Discovery Studio。
 - 2.4.1. 打開軟體,是否可看到下圖。勾選所有視窗如下。分別為 File, Protocols, Jobs, Tools, Parameters



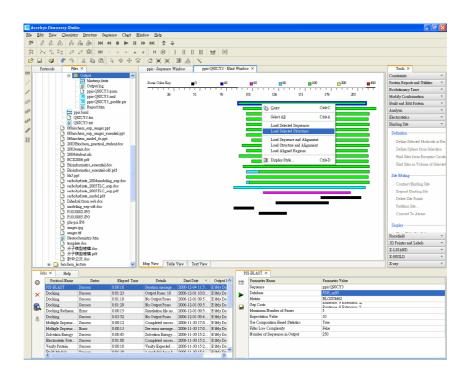
2.4.2. File \rightarrow New \rightarrow Sequence window



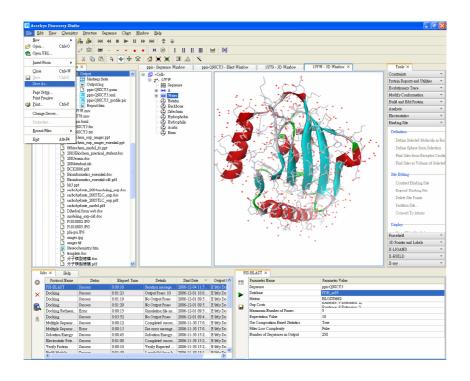
2.4.3. 貼上目標序列



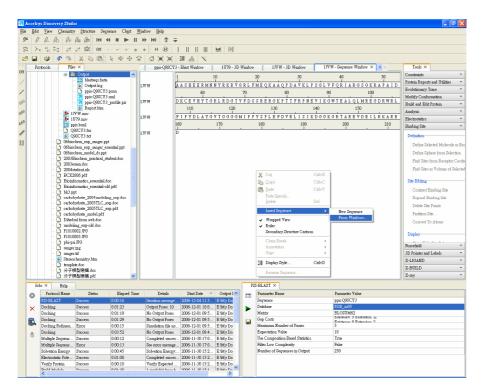
- 2.4.4. 點選 Protocols → Sequence → PsiBLAST,檢查修改 Parameters,點選綠色 三角執行。
- 2.4.5. 點選 Jobs → Success, 點選 Files → output
- 2.4.6. 右鍵點選模版 → Load selected structure



2.4.7. 如下圖,Sequence → show sequence



2.4.8. 空白處右鍵, insert sequence > From windows, 對話視窗選取目標序列。



- 2.4.9. 執行 Protocols → sequence alignment
- 2.4.10. 執行 Protocols → Build model
- 2.4.11. 依指示進行 minimization。
- 2.5. 對內建的 PDB 進行序列搜尋,比對,並進行分子模擬。依作業問題試著做做看。有問題請隨時舉手。

3. 參考文獻:

- 3.1. Anna Tramontano (2006) Protein structure prediction. Wiley-VCH
- 3.2. Höltje et al (2003) Molecular Modeling. Wiley-VCH
- 3.3. Current Protocol in Bioinformatics. Wiley online book.

M4 醣類分子模型

楊健志

碳水化合物的種類與構形非常複雜,原因是醣類的 hemiacetal 與另一醣類或醇類以醣苷鍵相連時,另一醣類不只一個位置具有羥基,且所形成的雙面角不像胜肽有特定的角度組合。Hemiacetal 形成時,會有兩種 anomer 生成,這對多醣體的構形有決定性的影響。利用分子模型組合醣類分子,可以清楚了解它的結構特徵。醣類修飾在其他生物分子,如蛋白質,脂質,賦予他們生物分子辨識特徵,增加水溶性等。

1. 器材

HGS Biochemistry Molecular Model, 5000 Bio-Organic Set (Maruzen, Japan) o

2. 實驗操作

2.1. 依據下圖 (Fischer projection) 建立一個葡萄糖 (P. 241, Lehninger)。

醣類模型

1
CHO
 1 CHO
 1 CHO

 1 CHO
 1 CHO

 1 CHO
 1 CHO

 1 CHO
 1 CHO

 1 CHO
 1 CHO

 1 CHO
 1 CHO

 1 CHO
 1 CHO

 1 CHO
 1 CHOH

 1 CHO
 1 CHOH

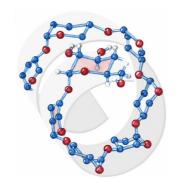
 1 CHO
 1 CHOH

 1 CHOH
 1 CHOH

 1 CHOH

- 2.2. 直鏈狀的六碳醣類很容易趨近於一個環狀,連結適當原子形成 hemiacetal。
- 2.3. 所完成的 pyranose 調整成 "椅形",有兩種可能形式, 4C_1 或 1C_4 合者較 為安定?

- 2.4. 試試看不同的 anomer 構形。
- 2.5. 以 α -1,4 linkage 連接兩個葡萄糖,與組員所的相連。觀察是否有親水性與疏水性的區域。有何重要性?



14 Biochemistry Laboratory

- 2.6. 以 β-1,4 linkage 連接兩個葡萄糖,與組員所的相連。
- 2.7. 試試 GlcNAc 及 GalNac。

M4 醣類分子模型 - 補充 TLC

楊健志

Oligosaccharide Analysis by Thin Layer Chromotography

β-Amylase (BA) (α-1,4-glucan maltohydrolase; EC. 3.2.1.2) catalyzes the removal of β-maltose residues sequentially from the non-reducing end of an α-1,4-glucan of variable chain length.

 β -Amylase is unique among other glycohydrolases that its catalysis proceeds with a repetitive manner, or multiple-attack mechanism. With this feature, β -amylase releases maltose effectively without dissociation from the rest of starch chain. This mechanism can be demonstrated when oligosaccharides are used as substrates. For example, when a maltose is removed from maltoheptaose, the maltopentaose still bound to the active site will be the substrate rather than a free maltoheptaose. Therefore, free maltopentaose will not be seen in the reaction mixture. If β -amylase works according to a multiple-chain mechanism, then free maltopentaose will be seen in the reaction system. We can trace the presence of maltopentaose by TLC.

To monitor the oligosaccharide released from multiple attack of β -amylase, a TLC method derived from Miyake et al was employed (Miyake et al., 2002).

Materials

Maltoheptose (G7)
TLC plate (Merck, Silica gel 60, 20*20 cm)
β-amylase (Sigma)
Development solution (H₂SO₄ (2%):ethanol =50:50)

Procedures

A 10 μ l mixture containing 1 μ l of maltoheptose, 1 μ l enzyme and 8 μ l sodium acetate buffer (50 mM, pH 5.4) was incubated at 37 °C for appropriate time (5, 10, 20, 30, 60 seconds, respectively). The reaction mixture (1 μ l) was then spotted on a TLC plate and developed by a solution of isopropanol:butanol:H₂O = 15:3.75:5. The spots was visualized by spraying a solution of H₂SO₄(2%):ethanol =50:50 and followed by heating at 180 °C for a few minutes.